NIH Research Festival 2025 — Bioinformatics Community Fair

- 10 AM – NIH Resources for Bioinformatics Data Analysis (Amy Stonelake, BTEP)
- 10:30 – Supercharge your Data Analysis with Biowulf (Antonio Ulloa, CIT)
- 11:00 – From Data to Discovery: Highlights from the Computational Genomics and Bioinformatics Branch (Daoud Meerzaman, CBIIT)
- 11:30 – Gen AI Community of Practice Tools and Training (Nick Weber, CIT and Dianne Babski, NLM)
- 12 noon – Bioinformatics Q & A and Lunch Break
- 1 PM – Gene Agent: An LLM Powered Tool for Gene Set Analysis (Zhiyong Lu, NLM/NCBI)
- 1:30 – AI in Clinical Imaging (Baris Turkbey, NCI/CCR)
- 2 – SCAssist: An AI Powered Workflow Assistant for Single-Cell Analysis (Vijay Nagarajan, NEI)
- 2:30 – Single-Cell and Spatial Transcriptomics SIG (Stefan Cordes, NHLBI)
- 3 – NIH Artificial Intelligence Interest Group (Ryan O'Neill, NHLBI)
- 3:30 - Bioinformatics Academic Offerings at FAES (Morgan Merriman, FAES)

NIH Resources for Bioinformatics Data Analysis

Amy Stonelake, Ph.D.
NCI CCR Bioinformatics Training and Education Program (BTEP), Program Manager

Sept 9, 2025
NIH Research Festival, Bioinformatics Community Fair

Bioinformatics Training and Education Program

NIH NATIONAL CANCER INSTITUTE
Center for Cancer Research

What we're going to cover.

- The NCI Center for Cancer Research (CCR) Bioinformatics Training and Education Program (BTEP) and NIH Bioinformatics Calendar
- NIH Library – Data Science, AI, Programming (Python/R), Data Management and Sharing
- NIH High-Performance Compute Cluster (Biowulf)
- NCI – Frederick-only based Compute Cluster (FRCE)
- NIH Cloud(s) – NIH STRIDES Program Cloud Lab, NCI CRDC (CGC, ISB-Gateway), HTAN, AnVIL (NHGRI), NIGMS Sandbox
- Production Workflows (NCI) – CCBR Github, CGC
- Software Licenses – Partek Flow, Qiagen Pathway Analysis, SnapGene, Qlucore, etc.
- Free NIH-wide License – Coursera Learning Platform (on pause til end of 2025/2026)
- NIH List Servs and Teams

NCI CCR Bioinformatics Training and Education Program (BTEP)

Began in 2012 to provide training in bioinformatics software

In 2024, 80+ classes with over 2,000 attendees

BTEP Team

Joe Wu, PhD, Trainer

Amy Stonelake, PhD, Program Manager

Alex Emmons, PhD, Trainer

Our goal is to enable scientists to understand and analyze their own experimental data by providing instruction and training in bioinformatics software, databases, analyses techniques, and emerging technologies.

Bioinformatics Training and Education Program Offerings

- AI in Biomedical Research Seminar Series
- Bioinformatics Bulletin (monthly)
- BTEP Class Catalog
- Class Documentation & Resource Pages
- Commercial Software (Partek Flow, etc)
- Distinguished Speakers Seminar Series
- Monthly Coding Clubs
- NIH Bioinformatics Calendar
- Office Hours – in-person and virtual
- Single-Cell and Spatial Transcriptomics Seminar Series
- Training events, classes, and webinars
- Video Archive

NATIONAL CANCER INSTITUTE
Center for Cancer Research

Home | Seminar Series | Announcements | Training | Resources | Contacts | My Account

Bioinformatics Training and Education Program

ncibtep@nih.gov

Bioinformatics Training & Education Program

Enabling scientists to understand and analyze their own experimental data by providing instruction and training in bioinformatics software, databases, analyses techniques, and emerging technologies.

Upcoming Classes & Events →

Provide feedback

Classes & Events

Browse Classes, Special Events, and Series Webinars.

Browse Class Schedule

Bioinformatics Resources

Class Documentation, Core Facilities, and Software.

Resources & Software

Bioinformatics Forums

Ask Questions about Bioinformatics Topics.

Questions & Answers

Video Archive

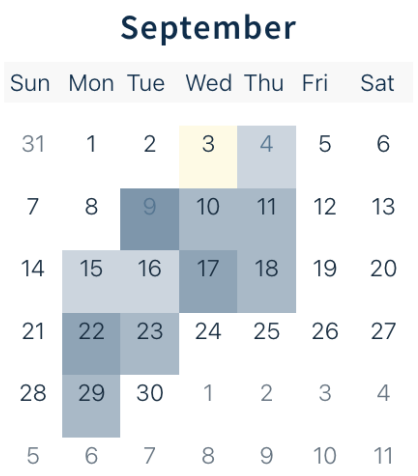
Class and Webinar Recordings and Transcripts.

Watch Videos

bioinformatics.ccr.cancer.gov/btep

Upcoming Classes & Events

NIH Bioinformatics Calendar



Tuesday
09

NIH Research Festival 2025

When: **Tue, Sep 09 - Fri, Sep 12, 2025 -9:00 am - 5:00 pm** Add to Calendar
Delivery: **In-Person**
Location: **Bethesda, Bldg 10**

Tuesday
09

Bioinformatics Community Fair at the NIH Research Festival

When: **Tue, Sep 09, 2025 - 10:00 am - 4:00 pm** Add to Calendar
Delivery: **In-Person**
Location: **Building 10 Library Training Room**
Presented By: **Amy Stonelake (BTEP), Antonio Ulloa (CIT), Baris Turkbey (CCR), Daoud Meerzaman (CBIIT), Morgan Merriman (FAES), Nick Weber (CIT), Ryan O'Neill (NHLBI), Stefan Cordes (NHLBI), Vijay Nagarajan (NEI), Zhiyong Lu (NCBI)**

Tuesday
09

Statistical Tests for Count Data

When: **Tue, Sep 09, 2025 - 12:00 pm - 1:00 pm** Add to Calendar
Delivery: **Hybrid**
Location: **Auditorium, Building 549, NCI at Frederick**
Presented By: **Duncan Donohue (Data Management Services Inc. a BRMI company)**

Wednesday
10

Rare Disease Informatics Workshop at the NIH Research Festival

When: **Wed, Sep 10, 2025 - 9:30 am - 11:00 am** Add to Calendar
Delivery: **In-Person**
Location: **NIH Bldg 10, FAES Classrooms 3 & 4**
Presented By: **Rare Disease Informatics SIG Members**

bioinformatics.ccr.cancer.gov/btep

BTEP Class Catalog (*new*)

Learning Level

- [ ] General Audience
- [x] Beginner
- [ ] Intermediate
- [ ] Advanced

Topic

- [ ] All Topics
- [ ] Artificial Intelligence (AI) (17)
- [ ] Cancer (3)
- [ ] Computing Resources (7)
- [ ] Data (6)
- [ ] Databases (7)
- [ ] Microbiome (1)
- [ ] Next Gen Sequencing (NGS) Methods (21)
- [ ] Omics (15)
- [ ] Programming (37)
- [ ] Software (16)
- [ ] Statistics (6)
- [ ] Technology (3)

Results found: 43

Computing Resources

2025

B4B Lesson 4: Working on Biowulf

Beginner

Computing Resources

2025

B4B Lesson 1: What is Biowulf?

Beginner

Data

2025

Documenting Analysis with Jupyter Lab

Beginner

Databases

2025

Pathway Analysis with Reactome

Beginner

Databases

2025

Functional Enrichment with DAVID

Beginner

Next Gen Sequencing (NGS) Methods

2025

Introduction to Gene Ontology and Pathway Analysis

Beginner

Next Gen Sequencing (NGS) Methods

2025

Performing Differential Expression Analysis for Bulk RNA Sequencing

Beginner

Next Gen Sequencing (NGS) Methods

2025

Differential Expression Analysis for Bulk RNA Sequencing: QC

Beginner

Next Gen Sequencing (NGS) Methods

2025

Aligning Next Generation Sequencing (NGS) Data to Genome

Beginner

Bioinformatics for Beginners (RNA-Seq)

- Module 1 – learning Unix on Biowulf
- Module 2 – bulk RNA-Seq data analysis (on Biowulf)
- Module 3 - pathway analysis (Reactome, DAVID, gene ontology)
- All class documentation and recordings available on BTEP website
- Will be offered again in 2026

Although we are a NCI/CCR resource…

We advertise our events throughout NIH

Our website is open to the world

We answer questions from ICs and outside of NIH

Video Archive is open and available

Class documentation and bioinformatics resources info are open

NIH Library Training Classes

- Artificial Intelligence (Prompt Engineering)
- Data Science
- MATLAB, SAS
- Programming (Python/R)
- Data Management and Sharing
- Other Library Services
- 3D Printing

nihlibrary.nih.gov

Options for Bioinformatics Analysis Work

LOCAL

HPC

CLOUD

What Bioinformatics Analysis can be done on your local machine?

Your NIH laptop or desktop

- Run Data Analysis Software (CLC Genome Workbench, SnapGene, Qiagen Pathway Analysis)
- Connect to HPC (high-performance cluster) via SSH or GUI
- Access the cloud(s)

How is the Cloud different from HPC?

- Cost based on usage
- Flexible compute power
- Analyze publicly available data

How is Cloud similar to HPC?

- Large compute resources
- Pre-installed software and tools
- Bioinformatics and data science analyses
- Analyze your own data

What is HPC (High-Performance Compute Cluster)?

Biowulf (NIH)

FRCE (Frederick)

Should I work on my local machine or Biowulf?

Why you should get to know Biowulf

Working on Biowulf – Things to Know

hpc.nih.gov

Minimal, fixed monthly cost

Thousands of scientific applications (software) available

Reference Data (NCBI db, BLAST db, genomic alignment data)

Login/head node

Learn some beginner Unix (Command Line)

Your home directory and your data directory

Start an interactive node (sinteractive)

Setting up swarm jobs (repetitive jobs)

Running batch jobs (scripts/programs)

Moving big data (Globus)

https://bioinformatics.ccr.cancer.gov/docs/resources-for-bioinformatics/Biowulf/

HPC on Demand GUI (RStudio, VSCode, JupyterNotebook, IGV, Matlab)

Frederick Research Computing Environment (FRCE)
NCI only

Available within NCI at no cost, optimized for Frederick researchers

250+ scientific applications installed

Next Gen Sequencing, structural biology, cryogenic electron microscopy, imaging and AI, text mining

Log in ssh or GUI, file transfers with Globus, batch and interactive jobs

https://ncifrederick.cancer.gov/staff/frce/

(Some) NIH Cloud Resources

- NIH STRIDES Initiative – NIH Cloud Lab accounts ($500)
- NCI Cancer Research Data Commons
- NHGRI AnVIL – platforms, workflows, data, community
- NIGMS Sandbox – data science learning modules

NIH STRIDES Iniative

- Accelerating Biomedical Research
- NIH Office of Data Science Strategy (ODSS)
- cloud.nih.gov
- NIH Cloud Lab

Publicly Available NCI Cloud Resources

- Cancer Research Data Commons (CRDC)
- Cancer Genomics Cloud/7Bridges/Velsera
- ISB Gateway in the Cloud/CGC
- Human Tumor Atlas Network (HTAN)

Cancer Research Data Commons

DATA COMMONS

Genomic Data Commons (GDC)

Share, analyze, and visualize harmonized genomic data, including TCGA, TARGET, and CPTAC.

Proteomic Data Commons (PDC)

Share, analyze, and visualize proteomic data, such as CPTAC and The International Cancer Proteogenome Consortium (ICPC).

Imaging Data Commons (IDC)

Share, analyze, and visualize multi-modal imaging data from both clinical and basic cancer research studies.

Integrated Canine Data Commons (ICDC)

Share data from canine clinical trials, including the PRE-medical Cancer Immunotherapy Network Canine Trials (PRECINCT) and the Comparative Oncology Program.

Clinical and Translational Data Commons (CTDC)

Store and share data from NCI-funded Clinical and Translational Studies.

General Commons (GC)

Hosting and sharing NCI data of multiple data types not a match for other CRDC Data Commons.

CORE RESOURCES AND SERVICES

Cancer Data Aggregator (CDA)

Enables users to query and connect data distributed across the CRDC for integrative analysis.

Data Commons Framework (DCF)

Provides secure user authentication and authorization and permanent digital object identifiers for data objects.

Seven Bridges Cancer Genomics Cloud developed by Velsera (SB-CGC)

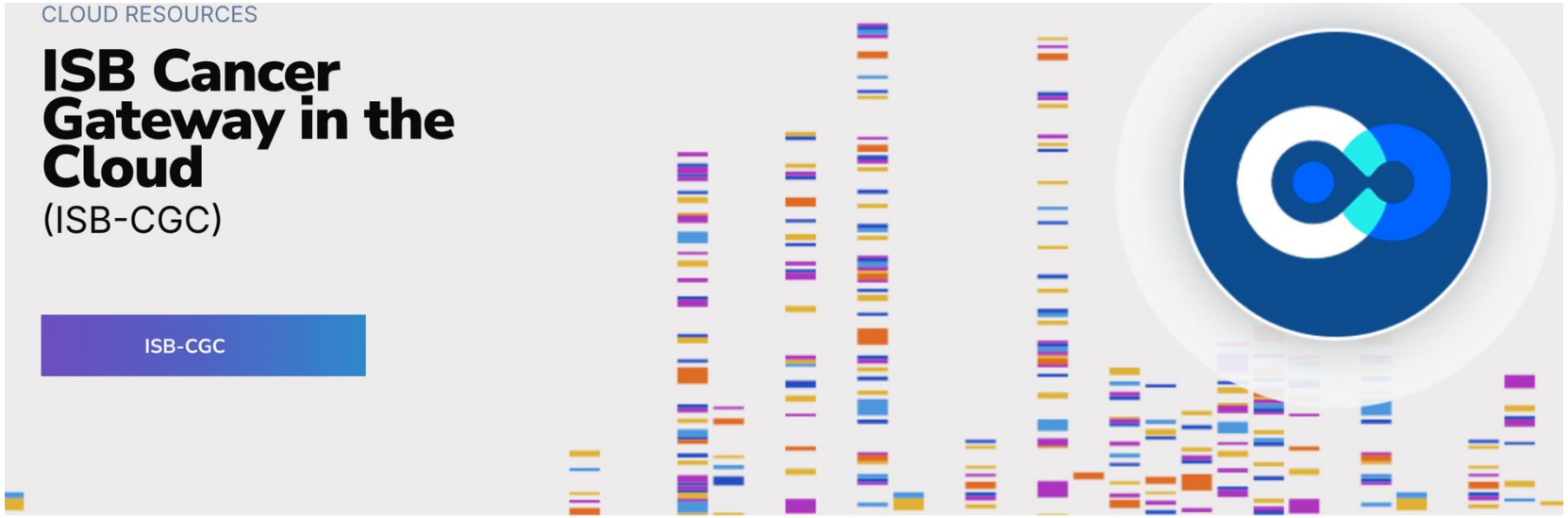
Enables users to query and connect data distributed across the CRDC for integrative analysis.

SevenBridges/Velsera/Cancer Genomics Cloud

Selected Workflows from NCI Cloud Resources

- \+ VARIANT CALLING
- \+ TRANSCRIPTOMICS
- \+ EPIGENOMICS
- \+ PROTEOMICS
- \+ IMAGING
- \+ MULTI-OMICS
- \+ SINGLE-CELL OMICS

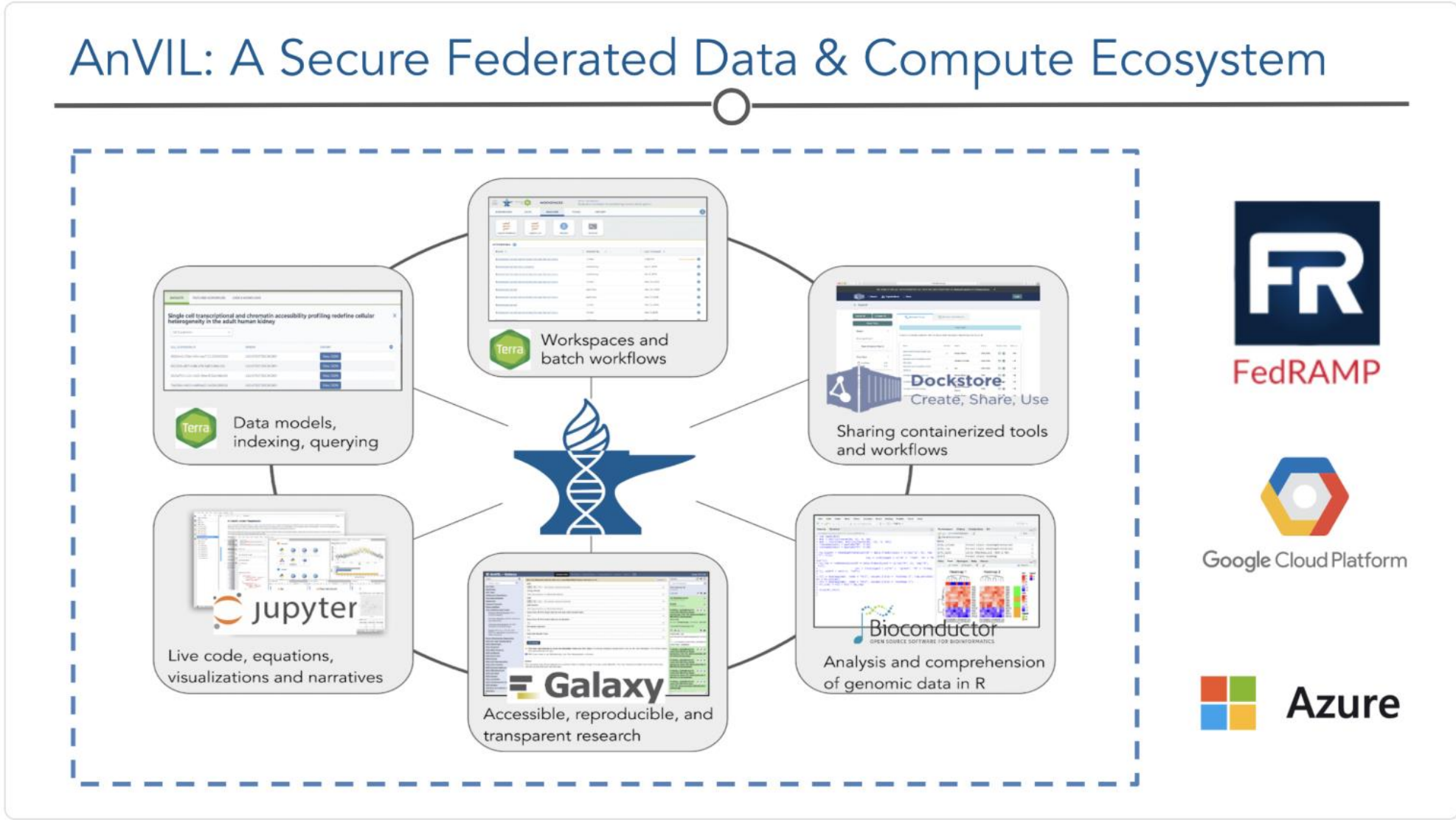
Institute for Systems Biology (ISB) Cancer Gateway in the Cloud



Overview

The ISB Cancer Gateway in the Cloud (ISB-CGC) was one of three original NCI Cancer Genomics Cloud Pilot programs that was funded from 2014 until 2025. For over ten (10) years, ISB-CGC enabled researchers to analyze cloud hosted data including TCGA and TARGET with their own tools as well as with a collection of powerful Google Cloud Platform native tools.

ISB continues to maintain and support the hosting of CRDC data within Google BigQuery tables, empowering users to search metadata and molecular cancer data from the Genomic Data Commons and other sources.



The NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space, or AnVIL, inverts the traditional model, providing a cloud environment for the analysis of large genomic and related datasets.

NHGRI Analysis Visualization and Informatics Lab-space (AnVIL)

anvilproject.org

Free trials available

Tools

Dockstore – create and share docker-based workflows

NCPI – interoperate with other NIH data commons

Bioconductor

Galaxy

Jupyter (python, R)

Data sources

Telomere-to-telomere genome (T2T)

1000G thousand genomes

Genotype-tissue expersion project (GTEx)

Workflows (pipelines)

CCBR Workflows

bioinformatics.ccr.cancer.gov/ccbr

NCI CCR Collaborative Bioinformatics Resource (CCBR) Pipelines

- CCBR pipelines (github/NIH)
- Free to use
- Consult CCBR in planning stage of experiment
- Don’t wait til you have data
- Submit help request for analysis project
- bioinformatics.ccr.cancer.gov/ccbr

Next Gen Seq Workflows/Pipelines
CCR Collaborative Bioinformatics Resource (CCBR)

github.com/ccbr

About Us

- 👋 Hi, we're the **@CCBR**, a group of bioinformatics analysts and engineers
- 📖 We build flexible, reproducible, workflows for next-generation sequencing data
- 💡 We collaborate with CCR PIs
- 📫 You can reach us at ccbr_pipeliner@mail.nih.gov
- 🏁 Check out our release history
- 🔗 Our Zenodo community

Available NGS Pipelines/Workflows

Here is a list of our prominent pipelines and their release schedule on BIOWULF:

Data Type	Pipeline Name	CLI* availability date	GUI* availability date
RNASeq[1]	RENEE snakemake	July 3rd 2023	July 14th 2023
WESSeq[2]	XAVIER snakemake	July 21th 2023	Sep 1st 2023
ATACSeq[3]	ASPEN snakemake	November 30th 2023	TBD
ChIPSeq[4]	CHAMPAGNE nextflow	October 15th 2023	TBD
CRISPRSeq[5]	CRISPIN nextflow	September 31st 2023	TBD
CUT&RunSeq[6]	CARLISLE snakemake	October 31st 2023	TBD
EV-Seq[10]	ESCAPE snakemake	March 26th, 2024	TBD
circRNASeq[7]	CHARLIE snakemake	*Jul 31st 2024*	TBD
scRNASeq[8]	SINCLAIR nextflow	*Sep 30th 2024*	TBD
WGSSeq[9]	LOGAN nextflow	*Sep 30th 2024*	TBD
spatialSeq[11]	SPENCER nextflow	TBD	TBD

Commercial Software Available (NCI)

Partek Flow – RNA-Seq, CITE-Seq, ATAC-Seq and Partek Genomics Suite

Qiagen Ingenuity Pathway Analysis (IPA) – from differential expression analysis (DEA) to pathways, biomarkers, drug targets

CLC Genomics Workbench

Qlucore Omics Explorer – RNA-Seq, visualizations, statistical analysis

Request access at service.cancer.gov (NCI)

SnapGene – molecular biology (alignment, PCR primers, cloning)

NIH GenAI Community of Practice

Nick Weber and Diane Babski (co-chairs)

nih-genaicommunity@groups.nih.gov

1200+ members

Holds monthly meetings and monthly Office Hours

AI Interest Group

Ryan O’Neill and Samar Samarjeet

ARTIFICIAL-INTELLIGENCE@LIST.NIH.GOV

AI Club –weekly Monday sessions in NIH Library

NIH AI Symposium (May, 2025) – one day, in-person, Bethesda campus

Sends ”Weekly Roundup” newsletter

list.nih.gov Interest Groups and Mailing Lists

ARTIFICIAL-INTELLIGENCE

BIOINFORMATICS-SIG-L

DATA-SCI ICE

SINGLECELLGENOMICS (AND SPATIAL TRANSCRIPTOMICS)L

SPATIALBIOLOGY@LIST.NIH.GOV

Email: ncibtep@nih.gov